Anno LV
Numero 2
Agosto 2006

Elettronica e telecomunicazioni

Editoriale

La sperimentazione Rai a Torino Finalista ad IBC2006

Realizzazione per mezzo di componenti a basso costo di un apparecchio dimostrativo per la misura della emissione umana di campi elettromagnetici a microonde

TV tridimensionale: sarà la volta buona?

Progettazione di un sistema fuzzy per l'annotazione automatica di contenuti multimediali

Edizione ottimizzata per la stampa.
La rivista è disponibile su web
alla URL www.crit.rai.it/eletel.htm

Anno LV
N° 2
Agosto 2006

Rivista quadrimestrale a cura della Rai

Indice

Editoriale

ing. Gianfranco **Barbieri**
Direttore di
"Elettronica e
Telecomunicazioni"

L'IBC *(International Broadcasting Convention*), che si tiene con frequenza annuale ad Amsterdam, costituisce uno degli appuntamenti più importanti, a livello mondiale, in cui vengono presentate le più recenti innovazioni tecnologiche conseguite nei comparti della radiodiffusione e della multimedialità.

Nell'edizione che si svolgerà quest'anno dal 7 al 12 Settembre, verranno attribuiti alcuni riconoscimenti all'innovazione (*IBC2006 Innovation Awards*) i quali hanno lo scopo di premiare l'eccellenza di applicazioni innovative delle tecnologie (hardware e software) che apportino un beneficio significativo in ambito creativo e commerciale.

Dei cinque riconoscimenti all'innovazione, il più prestigioso è il Judges' Award che premia applicazioni tecnologiche in cui convergano la creazione, la gestione e la distribuzione di contenuti in termini innovativi e che comportino benefici notevoli per gli utenti. I giudici, un gruppo internazionale di giornalisti della stampa specializzata, hanno concordato un ristretto elenco di finalisti candidati a questi riconoscimenti e la Rai è stata ritenuta meritevole di concorrere al prestigioso Judges' Award per la sperimentazione effettuata dal suo Centro Ricerche durante i XX Giochi Olimpici Invernali di Torino, di cui abbiamo dato dettagliata descrizione nel precedente numero della rivista.

Non si è ancora spenta l'eco dell'interesse suscitato dalla suddetta sperimentazione, che un nuovo fronte si sta aprendo sul versante dell'innovazione dei servizi. La visione tridimensionale è un traguardo ambizioso che i media dell'entertainment stanno da tempo inseguendo per rendere più appetibili alcune applicazioni destinate all'utenza consumer. I videogiochi, operando su un terreno tecnologicamente favorevole, costituiscono già una linea di frontiera consolidata; il settore della cinematografia, dal canto suo, per fronteggiare una crisi latente dovuta al calo degli spettatori è alla ricerca di soluzioni tecniche innovative e lo sviluppo di sistemi di proiezione per la visione stereoscopica è seguito con grande attenzione. L'articolo "TV tridimensionale: sarà la volta buona?" passa in rassegna una serie di applicazioni in cui la visualizzazione tridimensionale viene sfruttata per rendere più efficace il

servizio e descrive in sintesi lo stato dell'arte nel settore della TV stereoscopica compiendo una disamina delle problematiche ancora sul tappeto per lo sviluppo commerciale di questo servizio

Nel quadro delle attività volte allo sviluppo di tecnologie per l'automazione degli archivi si colloca la sperimentazione che il Centro Ricerche Rai sta compiendo, in collaborazione con ricercatori dell'Università di Torino, su sistemi per la caratterizzazione e di contenuti multimediali. Gli archivi multimediali sono una miniera di contenuti a cui attingere per le nuove produzioni e pertanto i metodi di reperimento ed accesso rapido al materiale archiviato diventano sempre più importanti; il problema del reperimento dei dati è strettamente legato a quello della documentazione. Come è noto, le operazioni di annotazione indicizzazione e predisposizione alla ricerca degli archivi audiovisivi richiedono oggi molte ore di lavoro manuale da parte dei documentatori. L'articolo "Progettazione di un sistema fuzzy per l'annotazione automatica di contenuti multimediali" presenta l'architettura di un sistema facente uso di tecniche di data mining basate sull'analisi del contenuto audiovisivo in unione con tecnologie del WEB semantico e sistemi intelligenti.

I lettori interessati ai problemi di inquinamento elettromagnetico troveranno in questo numero un articolo in cui viene descritta una soluzione a basso costo per la realizzazione di un apparecchio dimostrativo per la misura dell'emissione da parte del corpo umano di campi elettromagnetici a microonde. L'articolo evidenzia come la generazione di campi elettromagnetici non è soltanto artificiale ma anche naturale ed una insospettabile sorgente è ad esempio il corpo umano.

La sperimentazione Rai a Torino

Finalista ad IBC 2006

L'IBC (*International Broadcasting Convention*) è uno degli eventi più importanti a livello mondiale per quanto riguarda la presentazione delle innovazioni in campo tecnologico per la radiodiffusione e il multimedia e si tiene annualmente ad Amsterdam, a settembre.

Comprende una conferenza (quest'anno dal 7 all'11 settembre) ed una mostra (dall'8 al 12).

Il Centro Ricerche e Innovazione Tecnologica è presente con un contributo alla conferenza [1] nella sessione "HDTV We Get The Big Picture" del 7 settembre.

Quest'anno l'IBC ha introdotto quattro nuovi riconoscimenti per l'uso innovativo della tecnologia e un riconoscimento (Judges' Award) per un contributo particolarmente rilevante.

I giudici, un gruppo internazionale di giornalisti della stampa specializzata, hanno concordato un elenco dei finalisti per i vari riconoscimenti.

La sperimentazione promossa dalla Rai, ed organizzata dal Centro Ricerche, durante Torino 2006 [2], è stata ritenuta meritevole di competere come finalista al più prestigioso dei riconoscimenti, cioè il Judges' Award.

L'annuncio dei vincitori per i vari riconoscimenti è previsto in una apposita cerimonia (Award Ceremony) domenica 10 settembre.

Riferimenti Bibliografici

1. A. Morello, G. Alberico, M. Stroppiana: Rai HDTV and DVB-H Trials during the Turin Winter Olympics, IBC 2006 Conference, Amsterdam, September 2006

2. HDTV e TV Mobile, scintille di passione a Torino, Elettronica e Telecomunicazioni, anno 55, n. 1, aprile 2006

Realizzazione per mezzo di componenti a basso costo di un apparecchio dimostrativo per la misura della emissione umana di campi elettromagnetici a microonde

Maurizio **Tinti**
DEIS – Università di Bologna
Viale Risorgimento n°2
I-40136 Bologna
e-mail: maurizio.tinti@unibo.it

1. Introduzione

L'inquietudine della popolazione rispetto al rischio di danni alla salute provocati dai campi elettromagnetici che apparati ed impianti intenzionalmente o non intenzionalmente emettono è in questi ultimi anni fortemente aumentata e dalle cronache quotidiane si ha una chiara indicazione secondo la quale la preoccupazione è maggiore rispetto a quella per altri agenti fisici, per il fatto che il pubblico ha con essi una scarsa familiarità. Di conseguenza si genera una notevole divergenza fra le valutazioni dei rischi sanitari effettuate dalle organizzazioni scientifiche e la percezione degli stessi da parte dell'opinione pubblica. Si evidenzia quindi l'opportunità di dare ampia informazione al pubblico sulla natura dei campi elettromagnetici e sulla loro generazione, che non è soltanto artificiale ma anche naturale: una insospettabile sorgente di onde elettromagnetiche è, per esempio, l'uomo.

L'apparecchio dimostrativo descritto in questo articolo, valutando il campo emesso, in quanto corpi caldi, da una o più persone, può fornire un contributo all'informazione data al pubblico sulla generazione delle onde elettromagnetiche e può anche essere utilizzato per la misura delle radiazioni emesse dalle sorgenti naturali, delle quali occorre tenere conto nella fissazione dei limiti per le sorgenti artificiali.

Dell'apparecchio sono state costruite una versione per postazione fissa ed una portatile. La postazione fissa, dotata di filmato esplicativo, è stata installata nell'atrio della Villa Griffone a Pontecchio Marconi (Bologna), il luogo dove Guglielmo Marconi portò a compimento gli esperimenti sulla produzione e ricezione delle onde elettromagnetiche che aprirono le porte alle comunicazioni senza fili e che attualmente è sede di importanti attività di ricerca nel campo delle Radiocomunicazioni e delle Microonde. Gli studi marconiani hanno fornito anche un grande contributo alla nascita della Radioastronomia, disciplina che ci ha permesso di guardare l'Universo in un modo nuovo: proprio dalla strumentazione di tipo radioastronomico prende spunto questa realizzazione.

2. Descrizione dell'apparecchio

L'apparecchio che è stato realizzato rivela il campo elettromagnetico a microonde prodotto per la Legge di radiazione di Planck (figura 1) da una o più persone che si vengano a trovare davanti all'antenna di ricezione. Il pubblico può quindi sperimentare come sia esso stesso sorgente di emissione elettromagnetica.

È noto che i blocchi più critici degli apparati per la misura della radiazione elettromagnetica a microonde sono i ricevitori che, affinché i sistemi siano molto sensibili e possano misurare differenze di campo elettromagnetico molto piccole, devono essere a basso rumore proprio. Si rendono quindi necessari, per la realizzazione dei front-end dei ricevitori, ovvero delle sezioni a cui sono direttamente collegate le antenne riceventi, costosi amplificatori a basso rumore (LNA = *Low Noise Amplifier*), che nei casi di elevata professionalità possono essere raffreddati a basse temperature.

Da alcuni anni sono però presenti sul mercato "consumer" componenti a basso costo dedicati alla ricezione della TV satellitare detti LNB = *Low Noise Block* che, come sezione di front-end, impiegano un amplificatore a basso rumore. Per il nostro progetto si è allora pensato di impiegare come front-end di ricezione un LNB e come antenna un paraboloide per la ricezione della TV satellitare. All'apparecchio, infatti, non è affidata la specifica di misurare con precisione la temperatura di corpi caldi individuati tramite il fascio della sua antenna parabolica, specifica che richiederebbe una altissima stabilità del guadagno degli stadi di amplificazione, che non può essere soddisfatta da un LNB, ma deve unicamente avvertire ed interagire, con un notevole margine di sicurezza, ad una ΔT di sistema osservativo dell'ordine della decina di gradi (alla differenza di potenza ricevuta dall'antenna nel caso di presenza e di assenza di una persona davanti al fascio corrisponde, per la relazione di Nyquist, una differenza di temperatura "equivalente" a cui possono essere considerate persona e parete di fondo rispettivamente), quindi ben oltre i limiti teorici di frazioni di grado che comunque, come è stato analizzato, un LNB potrebbe garantire. L'applicazione non richiede inoltre la determinazione puntuale numerica delle temperature in gioco; ciò che è importante è raggiungere un rapporto segnale-disturbo S/N (del quale si parlerà nel seguito) tra "target" e fondo tale da garantire una buona visualizzazione dell'andamento nel tempo del campo.

Un LNB amplifica il segnale in ingresso che occupa una parte della banda Ku, assegnata alla TV via satellite nei limiti compresi fra 11 e 12,5 GHz, banda che, tramite un oscillatore locale interno,

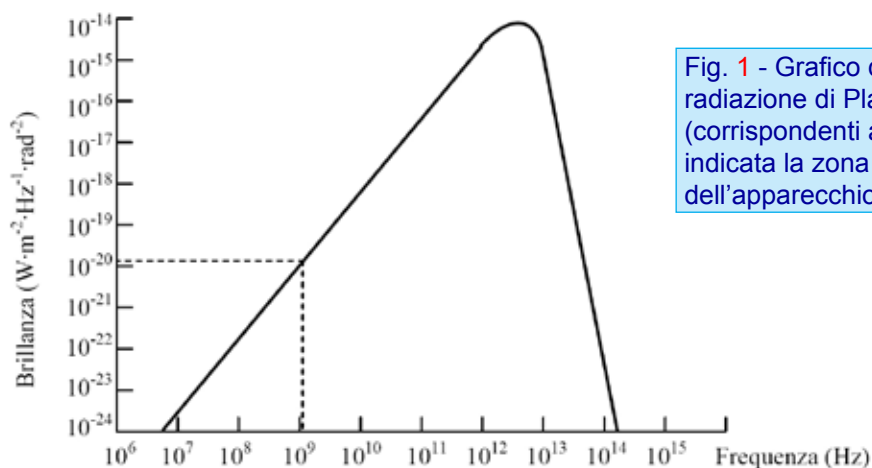



Fig. 1 - Grafico della Legge di radiazione di Planck a 37°C (corrispondenti a 310°K) con indicata la zona di funzionamento dell'apparecchio realizzato.

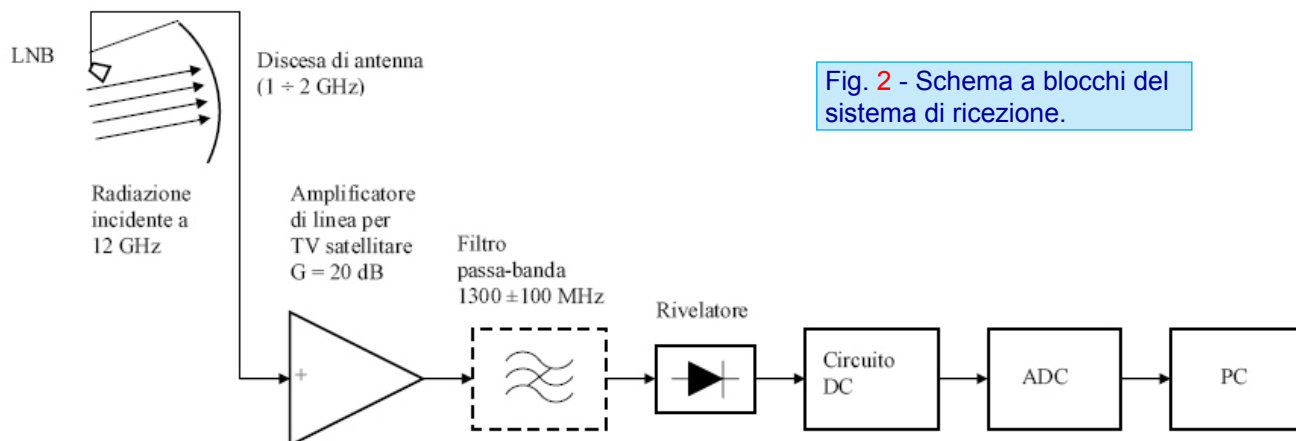



Fig. 2 - Schema a blocchi del sistema di ricezione.

viene "traslata" in una comoda media frequenza di uscita (IF) che, di norma, si estende fra 1 e 2 GHz. Il guadagno complessivo del blocco è di circa 60 dB su tutta la banda passante di circa 1 GHz. Queste notevoli prestazioni tecniche sono alla base della nostra decisione di usare tale componente; lo schema a blocchi in cui è inserito è rappresentato nella figura 2.

La misura della potenza di rumore ricevuta dall'LNB può avvenire "rivelando" il segnale di media frequenza per mezzo di un diodo, previa idonea amplificazione. Il livello ottimale di rumore che permette un buon funzionamento del diodo è dell'ordine di 0 dBm (1 mW), livello ottenuto tramite amplificazione a media frequenza. A tale incombenza provvede un amplificatore di linea per TV satellitare, che è anche questo un componente a basso prezzo presente nel mercato "consumer". Il suo guadagno vale 20 dB e la sua banda passante è circa 1 GHz (900 ÷ 2300 MHz), ovvero tutta la banda del segnale a media frequenza proveniente dal blocco LNB. Il circuito rivelatore, concepito secondo uno schema che prevede la compensazione in temperatura (si veda la figura 3), è stato montato all'interno del contenitore metallico dello stesso amplificatore di linea.

Nella versione per postazione fissa è stato interposto, fra amplificatore di linea e rivelatore (quest'ultimo montato, in questo caso, esternamente all'amplificatore, in un contenitore metallico a parte), un filtro passa-banda di tipo

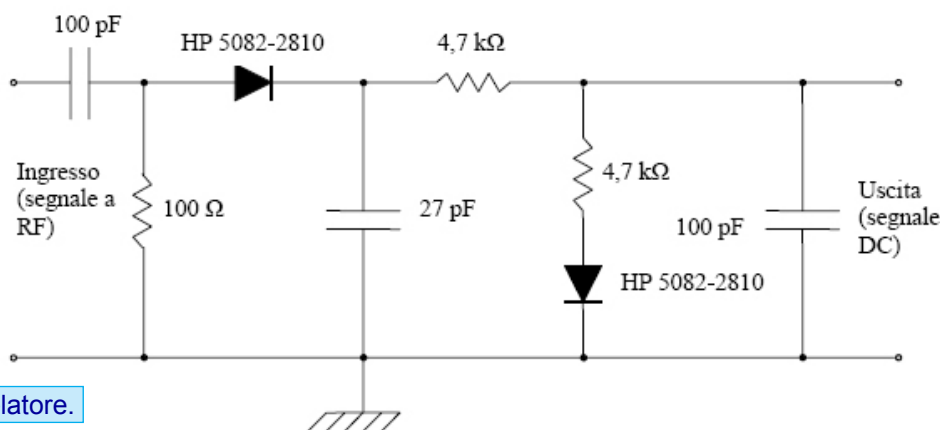
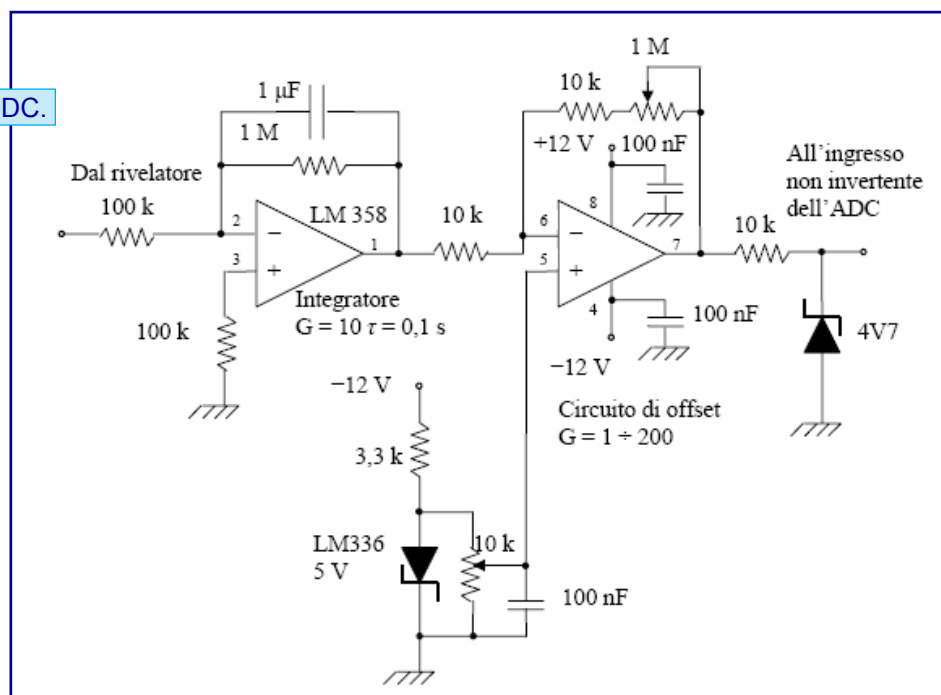



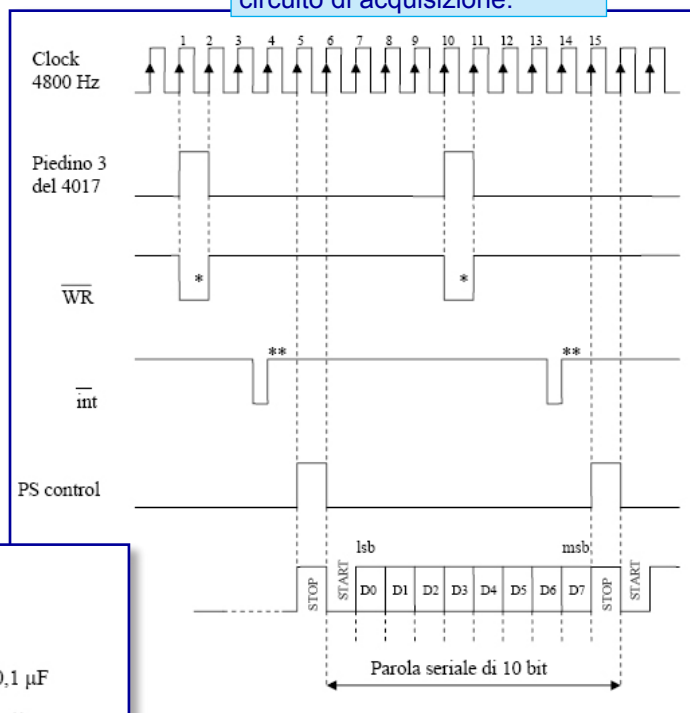
Fig. 3 -Circuito rivelatore.

Fig. 4 - Circuito DC.

interdigitato avente larghezza di banda di 200 MHz e frequenza di centro banda uguale a 1300 MHz. Tale limitazione di banda è stata operata per escludere il più possibile dalla grandezza misurata la presenza di segnali radio artificiali interferenti, a scapito però della sensibilità del ricevitore. Nella versione portatile il filtro non è stato inserito per motivi di compattezza della realizzazione.

L'uscita DC del rivelatore transita poi attraverso un circuito integratore ed un traslatore di livello (o circuito di off-set), realizzati utilizzando l'amplificatore operazionale doppio LM 358. Il circuito integratore compie una operazione di media delle fluttuazioni di ampiezza del segnale rivelato e quindi migliora la sensibilità del ricevitore. Il valore di compromesso per la costante di tempo dell'integratore è stata determinata sperimentalmente. Il traslatore di livello, posto a valle dell'integratore, è costituito da un circuito sommatore che permette di sottrarre, attraverso una regolazione manuale, un livello di tensione continua al segnale rivelato. Tale operazione di off-set risulta molto conveniente perché permette all'apparecchio di misurare unicamente la differenza fra la radiazione elettromagnetica di fondo e quella dell'oggetto verso cui è puntata l'antenna. Integratore e traslatore di livello, che costituiscono la sezione DC dell'apparecchio (rappresentata nella figura 4), effettuano anche una amplificazione del segnale in ingresso proveniente dal rivelatore, prima di inviarlo alla sezione di acquisizione dati.

La sezione di acquisizione dati (rappresentata nella figura 5) è stata realizzata, per garantire una elevata sicurezza di acquisizione e per ottenere precisione ed affidabilità elevate, interamente in modo hardware, anziché utilizzando componenti programmabili. La sezione è costituita da un convertitore analogico-digitale che, interfacciandosi alla porta seriale di un personal computer, permette, attraverso un opportuno software di acquisizione, la visualizzazione dell'andamento nel tempo della potenza di rumore ricevuta. La risoluzione del convertitore analogico-digitale è di 8 bit e la bit rate vale 4800 bit/s; i dati sono trasmessi a ciclo continuo, senza controllo di flusso, né hardware né software. Il campionamento e la conversione del segnale analogico sono affidati al circuito integrato ADC0804 che fornisce (in un tempo di circa 100 μs) il codice ad 8 bit paralleli corrispondente alla conversione di ogni campione di segnale acquisito. La successiva formazione della parola seriale richiede la conversione parallela-seriale con l'aggiunta dei bit di start e di stop. La prima operazione viene eseguita dallo

shift-register a 8 bit 4014: dopo che l'ADC ha reso disponibile il byte relativo ad un campione, avviene il suo caricamento parallelo nel registro a scorrimento. La seconda viene eseguita per mezzo del contatore decimale 4017, che fornisce anche il segnale di Write all'ADC. Sincrona con il segnale di clock fornito dal 4060 (contatore binario che, in questa applicazione, funge da generatore di clock quarzato), è disponibile al pin 11 del 4011 la parola seriale da inviare al PC. Il segnale di uscita viene infine adattato alla porta seriale dal circuito integrato driver MAX232. Le temporizzazioni della sezione di acquisizione dati sono rappresentate nella figura 6.

Fig. 6 - Temporizzazioni del circuito di acquisizione.

* inizio del processo di conversione da parte dell'ADC0804

** dati disponibili all'uscita dell'ADC0804

La parola seriale è formata da 10 bit: 1 Start Bit, 8 Data Bit, 1 Stop Bit. Non è presente il controllo di parità degli errori.

La velocità di trasmissione è 4800 bit/s quindi ogni parola viene trasmessa in circa 2,08 ms e vengono trasmessi 480 campioni al secondo.

Fig. 5 - Modulo di acquisizione.

L'alimentatore, che deve fornire una tensione di 12 V ed erogare una corrente di almeno 250 mA all'intero circuito (dei quali 150 mA occorrono per l'alimentazione, attraverso la discesa di antenna, del blocco LNB), è basato sul regolatore di tensione 7812 ed è mostrato nella figura 7. La tensione di 5 V, necessaria per l'alimentazione dei circuiti logici, è ottenuta per mezzo di un regolatore 7805, mentre la tensione di −12 V, necessaria per l'alimentazione simmetrica degli amplificatori operazionali, è ottenuta per mezzo del convertitore ICL7660A.

Le sezioni DC, di acquisizione e l'alimentatore (trasformatore escluso) sono stati montati su piastra millefori, ed a quest'ultima è stata fissata la sezione a radiofrequenza. Il tutto è stato alloggiato in un contenitore plastico standard.

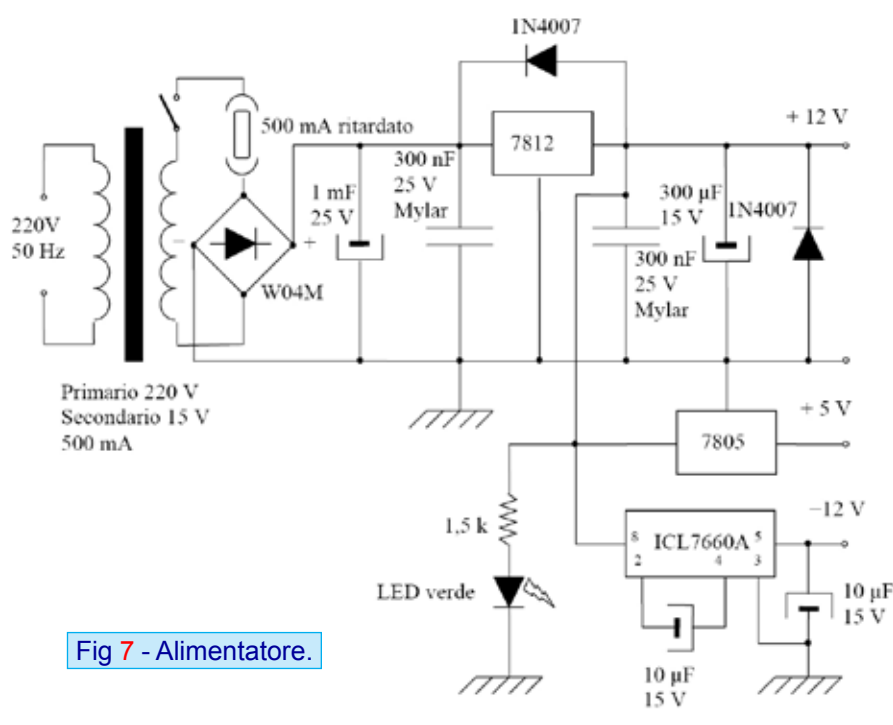
Nella fase di messa a punto del ricevitore viene regolato il potenziometro di offset e quello del guadagno dell'amplificatore DC in modo che l'intervallo di variabilità dei valori di tensione DC all'ingresso del convertitore analogico-digitale sia compreso all'interno del suo intervallo di lavoro (0 ÷ 5 V).

Fig 7 - Alimentatore.

3. Verifica delle prestazioni del sistema

Realizzate le prima citate condizioni operative iniziali, si è posto il problema della verifica delle prestazioni radio-elettriche del sistema. I calcoli teorici relativi alla minima temperatura detectabile da questo apparecchio, che è risultata essere di 0,223 °K, mettono in evidenza la quasi risibile facilità con cui un corpo caldo operante a 37 °C rispetto ai 20 ÷ 25 °C della temperatura dell'ambiente circostante possa essere individuato e selezionato dal fascio della antenna parabolica. Una attenta considerazione sulle temperature effettivamente in gioco all'ingresso del ricevitore, tendono però sfortunatamente a peggiorare il rapporto S/N: si pensi ai lobi secondari dell'antenna che "vedono", ancorché ridotta, la temperatura di tutti gli oggetti circostanti, la temperatura di rumore dello stesso "front end" del ricevitore, le perdite dei cavetti coassiali di collegamento, la modesta efficienza dell'antenna e il suo "effetto diluizione" che si verifica quando il corpo caldo da misurare è inferiore al fascio primario dell'antenna, che è stato calcolato essere uguale a 3°.

Una stima delle temperature di ingresso del sistema ricevente, considerati tutti i contributi citati, permette di valutare il rapporto S/N di lavoro, modesto ma ancora superiore al limite teorico; pertanto si può garantire un adeguato funzionamento dell'apparato in discussione.

Una serie di misure di sorgenti naturali (Terra, Sole, Cielo freddo), effettuate sul prato antistante Villa Griffone, hanno dimostrato la pressoché perfetta corrispondenza tra i rapporti di potenza misurati e quelli attesi a seguito di

stime o dati teorici. Non abbiamo notato la presenza di segnali radio interferenti che ci hanno invece impedito ogni sorta di misurazione in un laboratorio CNR a Bologna centro. È stato verificato con l'analizzatore di spettro che si trattava di uno o più ponti ripetitori televisivi locali operanti esattamente sulla nostra banda di ricezione.

4. Installazione del sistema

Durante l'installazione del sistema sono state effettuate prove per il posizionamento migliore dell'antenna nell'atrio della Villa Griffone (figura 8). Nei pressi della parete accanto alla quale vi è spazio sufficiente per ospitare il sistema si è posizionata l'antenna a varie altezze e secondo numerosi orientamenti. Per ogni posizione (in assenza di persone davanti al fascio di antenna) si è rilevata l'indicazione (relativa) dello strumento di misura. La procedura è stata ripetuta puntando l'antenna su una persona che sta entrando nella Villa.

Il posizionamento migliore per l'antenna è quello per cui:

1) la potenza di rumore misurata risulta essere minima in assenza del soggetto. Ciò si ha se, in assenza di segnali radio artificiali interferenti, l'antenna vede una parete che si trova alla temperatura più bassa possibile (in presenza di segnali radio artificiali interferenti la condizione di minima potenza di rumore misurata può essere soddisfatta per puntamenti dell'antenna diversi da quello verso la parete a temperatura minima);

Fig 8 - Stazione dimostrativa installata nell'atrio della Villa Griffone a Pontecchio Marconi.

2) la potenza di rumore misurata risulta essere massima in presenza del soggetto. Ciò si ha se l'antenna punta il soggetto in modo che quest'ultimo occupi la parte più grande possibile dell'area sottesa dal fascio dell'antenna e se il soggetto si trova sufficientemente vicino ad essa (qualche metro).

A queste condizioni infatti è massima la differenza fra il rumore misurato in assenza ed in presenza del soggetto e quindi, effettuato l'offset dell'apparecchio, sono più improbabili false rivelazioni. A queste condizioni si aggiunge la:

3) l'antenna deve vedere la parete che subisce le più piccole variazioni di temperatura nell'intera giornata.

In questo modo la variazione di rumore che il radiometro misura, in assenza del soggetto, durante l'intera giornata è quella minima e quindi non dovrebbero essere necessarie ripetute regolazioni dell'offset. La condizione 1) è risultata essere verificata posizionando l'antenna a circa tre metri di altezza rispetto al pavimento ed orientandola verso il basso in modo da farla puntare obliquamente verso il pavimento. La condizione 2) è risultata essere verificata orientando l'antenna in modo da farla puntare nella zona del pavimento in cui si presume transiterà la persona da rivelare (circa al centro dell'atrio della Villa). La condizione 3) è stata soddisfatta posizionando l'antenna in modo che puntasse verso una zona dell'atrio in cui non batte mai il Sole.

5. Conclusioni

La scarsa conoscenza della natura dei campi elettromagnetici da parte del pubblico può portare ad una percezione del rischio che deriva dalla loro emissione molto maggiore rispetto a quello determinato dagli studi scientifici nel settore. Specifici apparecchi, come quello descritto in questo articolo, possono rendersi utili per una efficace comunicazione di queste problematiche e per agevolare la familiarizzazione del pubblico verso la materia campi elettromagnetici, per la cui comprensione è richiesta una preparazione specifica.

Bibliografia

1. G. Tomassetti, Esperimenti di Radioastronomia amatoriale, Istituto di Radioastronomia - Consiglio Nazionale delle Ricerche - Bologna, 2000

2. John D. Kraus, Radio Astronomy, Mc Graw - Hill Book Company, New York, 1966

3. H. Eriksson R.W. Waugh, A Temperature Compensated Linear Diode Detector, Design Tip Agilent

4. F. Falcinelli, Amplificatore DC-integratore di post-rivelazione con interfaccia per acquisizione dati

Ringraziamenti

Ringrazio sentitamente il Prof. Gabriele Falciasecca che con lungimiranza ha intuito il successo che questo apparecchio avrebbe riscosso e mi ha incoraggiato nel compimento di questo lavoro. Un particolare ringraziamento al Dott. Goliardo Tomassetti, ideatore dell'apparecchio, che con grande disponibilità mi ha pazientemente seguito nel corso di tutta la realizzazione. Ringrazio inoltre il Sig. Luigi Amaduzzi per i preziosi consigli sulla parte di acquisizione del segnale.

TV tridimensionale:

sarà la volta buona?

ing. Mario Muratori
Rai - Centro Ricerche e Innovazione Tecnologica

1. Arriva il 3D

Forse sì.

Intanto bisogna rilevare che la visualizzazione tridimensionale, e la stereoscopia in particolare, è da tempo utilizzata in quei settori professionali dove dà la possibilità di ottenere risultati altrimenti impossibili o di svolgere funzioni in modo efficiente ed efficace. Per esempio, nel campo della cartografia è normale ricorrere all'esame stereoscopico di materiale fotogrammetrico per il riconoscimento di particolari del terreno altrimenti analizzabili solo con sopralluoghi. Nel campo della progettazione, i CAD tridimensionali sono ormai diffusissimi e, quando i modelli sono di una certa complessità, la visualizzazione tridimensionale si impone come strumento indispensabile; in particolare, nel campo automobilistico la visualizzazione tridimensionale è utilizzata come strumento per l'analisi e la presentazione di modelli, per esempio relativi a carrozzerie o interni, prima della loro realizzazione prototipale.

Sommario
Attualmente vi è una forte concorrenza tra i vari settori dell'industria dello spettacolo – in questo articolo sono considerati i videogiochi, il cinema, la televisione e l'home video – ed è sensibile lo sforzo di ciascun settore di eccellere rispetto agli altri, anche dal punto di vista tecnologico, al fine di attirare una maggiore quantità di spettatori.
In questa situazione, la visualizzazione tridimensionale è considerata uno strumento importante poiché è in grado di suscitare una notevole sensazione di immersività e quindi un maggiore appagamento dello spettatore.
Nei vari settori si usano tecnologie di visualizzazione tridimensionale differenti e il grado di utilizzazione delle stesse varia notevolmente. In questo articolo si effettua una disamina dei sistemi utilizzati e delle tendenze in atto.

Nel campo aeronautico, in particolare militare, si stanno adottando sistemi di visualizzazione tridimensionale per proporre agli operatori, in modo ergonomico ed efficace, la situazione operativa, che in genere si sviluppa nel mondo tridimensionale, arricchendone il modello sintetico con gli opportuni parametri operativi. I robot per operazioni chirurgiche, la cui adozione si sta diffondendo, sono spesso corredati da apparati di visualizzazione tridimensionale per rendere più realistica la rappresentazione delle parti anatomiche interessate all'operazione e quindi più precisa l'azione del chirurgo.

In queste tipologie di utilizzo, le tecniche di visualizzazione tridimensionale si possono considerare assodate, ancorché non generalizzate, poiché giustificate dai risultati che permettono anche a fronte di costi, a volte, ragguardevoli.

Nel campo dello spettacolo e del divertimento entrano in gioco altri fattori, primo tra i quali la necessità dei produttori di offrire spettacoli sempre più avvincenti per richiamare pubblico e, nello stesso tempo, di limitare al massimo i costi per gli utenti. Per queste ragioni, pur esistendo da tempo, le tecniche di visualizzazione tridimensionale non sono state finora molto utilizzate nel campo dello spettacolo, salvo alcuni sporadici tentativi. Tuttavia sembrerebbe che le cose stiano cambiando, in parte perché la tecnologia odierna permette di ridurre i costi e in parte perché l'utenza diventa sempre più esigente e la concorrenza tra i vari media sempre più agguerrita: il mondo dei videogiochi è un fortissimo concorrente della televisione e del cinema, per cui per questi ultimi diventa un riferimento, possibilmente da superare, anche dal punto di vista tecnologico.

2. Videogiochi

Sembra, infatti, che nel campo dei videogiochi la visualizzazione tridimensionale sia un fatto acquisito, anche se limitato a quella nicchia di utenti disposti ad equipaggiarsi di elaboratore di elevate prestazioni e sistemi di visualizzazione adatti, a partire dagli occhiali shutter per finire agli elmetti con visualizzazione binoculare (HMD – Head Mount Display).

Fig. 1 - Esempi di HMD e vista dall'interno

I videogiochi si prestano particolarmente bene all'utilizzazione delle tecnologie tridimensionali: l'ambiente in cui si svolge il gioco è completamente virtuale e non solo è relativamente semplice da generare, ma permette la complessità di gioco e la spettacolarità ambientale che il videogiocatore odierno ricerca. La visualizzazione del prodotto richiede elaboratori di capacità computazionali relativamente elevate, ma comunque non superiori a quelle offerte dai personal computer di fascia medio-alta equipaggiati con schede grafiche avanzate che si trovano comunemente sul mercato. Gli HDM di qualità "professionale" (p.es. risoluzione 1280x1024) hanno costi piuttosto elevati, dell'ordine di grandezza delle decine di migliaia di euro, mentre gli HDM di livello "consumer" (p.es. risoluzione 800x600) e i sistemi basati su occhiali shutter hanno costi più accessibili, anche meno di 1000 €. Perciò il costo necessario per acquisire un sistema elaborativo adatto per videogiochi tridimensionali è alla portata di moltissime persone.

3. Cinema

Per fronteggiare la diminuzione di spettatori, l'industria cinematografica si sta indirizzando verso tecnologie di visualizzazione tridimensionali [1, 2], anche sfruttando il fatto che il prodotto viene fruito nei cinema, dove la visione tridimensionale può essere sfruttata efficacemente per aumentare la sensazione di immersività dello spettatore, suscitando in esso forti emozioni – e di conseguenza la soddisfazione dell'utente pagante.

La tecnologia di visualizzazione utilizzata in cinematografia è essenzialmente quella stereoscopica, in particolare con utilizzo dell'anàglifo e della luce polarizzata. Tale tecnologia richiede l'uso di occhiali speciali di tipo passivo, in genere composti da una montatura di cartoncino e lenti costituite da pellicola plastica, che oltre ad essere poco costosi possono anche fungere da gadget. La proiezione dei film in locali di caratteristiche controllabili, in particolare quelle geometriche e illuminotecniche, permette di ottimizzare i parametri di produzione al fine di massimizzare

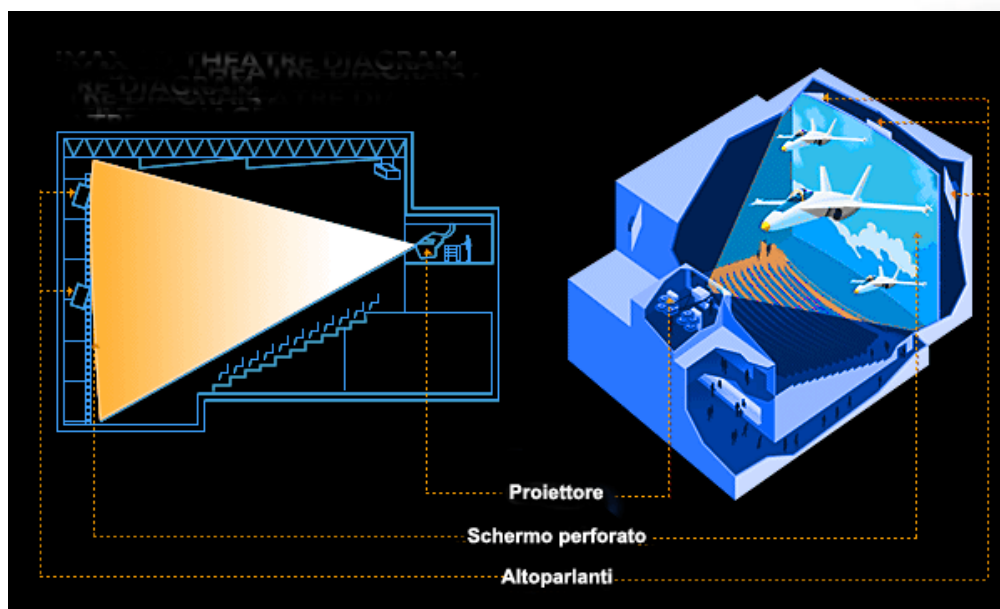



Fig. 2 - Sala IMAX stereoscopica e spettatori

l'effetto sullo spettatore. Inoltre la proiezione nei cinema, che sono luoghi chiusi, permette di gestire facilmente la distribuzione degli occhiali ed eventualmente la loro restituzione, opportuna qualora si usino oggetti di pregio.

Un esempio particolare è il sistema IMAX che associa uno schermo di grandi dimensioni, circa 600 m^2, con la possibilità, nella versione IMAX 3D, di utilizzare varie tecniche stereoscopiche. Lo schermo di grandi dimensioni suscita una sensazione di immersività molto coinvolgente poiché sfrutta la visione periferica; inoltre tale sensazione viene incrementata dalla visione tridimensionale delle scene. Il risultato sembra avere le potenzialità per essere piuttosto trascinante [3, 4].

Ovviamente uno dei problemi è la produzione di film stereoscopici in numero sufficiente ad alimentare le sale idonee con una ricchezza di titoli tale da attirare lo spettatore, nonché giustificare gli investimenti necessari per costruire nuove sale, o quanto meno adattare quelle esistenti, con l'equipaggiamento apposito per la stereoscopia. I teatri IMAX 3D, ad esempio, propongono una produzione recentissima di film stereoscopici realizzata specificamente, ma a causa della ridotta diffusione del formato, in Italia vi sono sale IMAX 3D solo a Riccione e a Taranto, e solo quest'ultima pubblicizza la proiezione di materiale stereo, non si può ancora sostenere che sia un tipo di spettacolo da considerarsi "normale".

Fig. 3 - Cinepresa stereoscopica in formato IMAX 3D.

Una soluzione parziale viene dai sistemi capaci di estrarre le informazioni di profondità dall'analisi automatica o semiautomatica delle scene, producendo una coppia stereoscopica[Nota 1].

Tali sistemi trovano una giustificazione economica nella trasformazione in tridimensionale di prodotti già esistenti, ma per le nuove produzioni non sono economicamente competitivi nei confronti della realizzazione effettuata direttamente con tecniche tridimensionali. D'altro canto, l'uso delle tecniche stereoscopiche già in sede di ripresa e di produzione virtuale dà risultati tecnicamente migliori e lascia maggiore libertà creativa.

In genere si preferisce effettuare le riprese con apparati elettronici digitali di elevata qualità [5, 6], sia per abbattere i costi, sia per opportunità tecnica. Il prodotto ottenuto è poi facilmente integrato con quello generato in sede di post-produzione - quali effetti speciali, scenografie e soggetti virtuali, ecc. - dove si applicano alcuni accorgimenti specifici per tenere conto della composizione tridimensionale della scena, ma che non richiede costi di molto superiori alla tradizionale post-produzione bidimensionale.

Quest'ultimo fattore, soprattutto, porta alla conclusione che nel prossimo futuro si potrebbe assistere ad una elevata produzione di film d'animazione tridimensionali.

Nota 1 - In particolare ci si riferisce al processo chiamato "Dimensionalization" proposto dalla In-Three Inc. - www.in-three.com.

Si noti che, secondo alcune fonti, il cinema digitale è l'arma cui deve ricorrere l'industria cinematografica per sollevarsi dalla crisi, in particolare per la sua maggiore efficacia nei confronti della pirateria.

Tuttavia, questa tecnologia è interessante anche perché si integra più facilmente con la produzione e post-produzione effettuata in "computer grafica", quindi con effetti speciali molto complessi, e permette di adottare facilmente le tecnologie di realizzazione, lavorazione e visualizzazione stereoscopiche al fine di incrementare ulteriormente l'interesse per il prodotto finale [7].

4. Televisione

Il mondo della televisione, dopo avere introdotto i sistemi di registrazione (DVD) e di trasmissione (DVB-S, DVB-T, DVB-H) digitali, attualmente è alle prese con l'introduzione dell'alta definizione (HDTV – High Definition TV).

Ci vorranno alcuni anni perché il mercato ed il sistema tecnologico nel suo insieme - cioè la filiera produttiva e distributiva - si assestino e che quindi si possano presentare con successo nuove proposte, ma si inizia già da ora ad analizzare le implicazioni – tecniche, produttive e di mercato – dell'introduzione di nuovi sistemi televisivi a visione tridimensionale che sembrano essere la prossima frontiera.

Attualmente, l'olografia sembra essere una tecnologia ancora troppo acerba [8, 9, 10, 11]. Di conseguenza, maggiore attenzione viene rivolta ai sistemi stereoscopici, tanto più che sono inerentemente compatibili con i sistemi bidimensionali: basta infatti visualizzare un canale solo della coppia stereo.

In linea di principio, le problematiche della produzione e della post-produzione non sono diverse da quelle relative alla cinematografia, seppure con importanti differenze, alcune in positivo e altre in negativo. Non pare abbia molto senso adottare definizioni superiori a quella della TV ad alta definizione (1920 pixel/1080 righe), il che permette quindi di utilizzare sistemi e apparati già a disposizione per quest'ultima, in contrasto col cinema digitale per il quale si adottano definizioni uguali o superiori a quella nota come 2K (2048 pixel/riga). D'altro canto, una caratteristica peculiare della produzione televisiva è costituita dalla programmazione di informazione e cronaca, caratterizzata da sistemi di ripresa maneggevoli e robusti per permettere un'elevata mobilità, mentre l'ambiente di ripresa cinematografica è più controllato e vi si possono utilizzare anche sistemi complessi e relativamente delicati.

Uno dei principali fattori che tradizionalmente si è rivelato limitante per lo sviluppo della stereoscopia televisiva è la struttura di distribuzione e diffusione, in particolare a causa della insufficiente capacità di trasporto[Nota 2]. Con i sistemi digitali introdotti nell'ultimo decennio si è superata questa limitazione ed attualmente vi è una maggiore disponibilità di canali. Con lo standard DVB-S si è normalizzata la diffusione televisiva diretta via satellite, il cui successo è stato tale da spingere le aziende di gestione di satelliti a lanciarne un numero notevole, tanto che attualmente si rileva un surplus di banda disponibile. La diffusione digitale terrestre (DTT), normalizzata con lo standard DVB-T, permette di veicolare più canali televisivi[Nota 3] nella stessa banda occupata precedentemente da un solo canale analogico, di fatto aumentando le capacità trasmissive in termini di canali televisivi. Anche in questo caso, in un'ottica di abbandono dei sistemi trasmissivi analogici, si otterrà una capacità trasmissiva superiore a quanto disponibile oggigiorno, probabilmente solo parzialmente

Nota 2 - Con particolare riferimento al numero di canali televisivi possibili in reti estese sul territorio nazionale.

Nota 3 - Attualmente la RAI adotta un bouquet di 5 canali.

Fig. 4 - Schermo HDTV 42” autostereoscopico a microlenti prodotto dalla Philips

Fig. 5 - Coppia di proiettori montati su apposito supporto con filtri polarizzatori per proiezione stereoscopica

occupata dalla maggiore capacità richiesta dalla HDTV di prossima introduzione.

In aggiunta ai nuovi sistemi trasmissivi, si stanno sviluppando nuovi e più avanzati metodi di codifica (p.es. MPEG4-AVC) che permetteranno una ulteriore riduzione della capacità richiesta per la trasmissione di segnali televisivi digitali e quindi maggiori capacità trasmissive sui canali esistenti (che hanno canalizzazioni fissate da normative internazionali).

Complessivamente, quindi, la situazione si sta evolvendo verso la possibilità - tecnicamente fattibile e economicamente giustificabile - di inserire anche coppie stereoscopiche, eventualmente HDTV, nei bouquet delle trasmissioni digitali.

L’ultimo apparato della catena, quello di visualizzazione, rimane invece il punto dolente di ogni tecnica stereoscopica.

Infatti l’utenza televisiva di elezione è quella domestica, caratterizzata da condizioni di visione tutt’altro che controllabili: si hanno infatti dimensioni degli schermi diversissime, distanze di visione non ottimali ma obbligate dalle dimensioni dei locali di abitazione, illuminazione dell’ambiente di visione altrettanto non controllata e molto differenziata. Altre carat-

teristiche salienti sono la visione multiutente e, non ultima, la necessità che gli apparati abbiano costi ridotti.

Con queste situazioni, gli apparati di visualizzazione più indicati sarebbero i monitor autostereoscopici a microlenti, i cui pregi principali consistono nella visione tridimensionale multiutente e senza ausilii (è infatti ottenuta dal tipo di costruzione dello schermo) e le buone caratteristiche illuminotecniche, di ingombro e di consumo paragonabili ai comuni monitor TFT dai quali derivano. Purtroppo in genere hanno l'inconveniente di non permettere una buona visione del segnale bidimensionale, di avere una risoluzione orizzontale[Nota 4] limitata, di richiedere elaborazioni complesse se alimentati da una coppia stereoscopica[Nota 5] e di avere ancora un prezzo elevato. Inoltre, attualmente si trovano in commercio in genere monitor di questo tipo associati a software e schede specifiche da installare su personal computer, ma non ancora pronti per l'uso televisivo [12, 13, 14, 15].

Nota 4 - Infatti, per offrire una visione stereoscopica in un angolo di visione accettabile (p.es. 160°) si devono utilizzare 7 o 9 "viste", ognuna delle quali richiede un pixel. Pertanto la risoluzione orizzontale degli schermi base, normalmente del tipo TFT, viene ridotta significativamente, anche se si utilizzano più righe riducendo anche la risoluzione verticale.

Nota 5 - La coppia stereoscopica veicola due "viste". Quelle mancanti per arrivare al numero accettabile di 7 o 9 devono essere ricostruite al ricevitore. A tutte le viste, originali e ricostruite, si devono poi applicare filtri, in genere bidimensionali e piuttosto complessi, e sotto campionamenti ottimizzati per ciascun tipo di schermo (in dipendenza dal tipo di microlente adottata, dal numero di viste, dalla risoluzione del monitor, ecc.).

Fig. 6 - Schermo a retroproiezione di grande formato. I sistemi costituiti da una coppia di proiettori e una lente frontale opportuna sono adatti per la proiezione stereoscopica.

Mentre i monitor autostereoscopici a microlenti sono apparati specificamente costruiti per la stereoscopia, i sistemi basati su proiettori utilizzano tecnologie correnti, ancorché possano essere anche ad alta definizione e quindi pronti per l'eventuale HDTV stereoscopica, godendo dei benefici economici del caso. In questi sistemi, che siano a proiezione diretta o a retroproiezione, i due canali costituenti la coppia stereoscopica vengono proiettati contemporaneamente e la separazione avviene sfruttando le caratteristiche di polarizzazione o di frequenza della luce[Nota 6].

In ambedue i casi, l'utente ha la necessità di indossare speciali occhiali muniti di lenti adatte per separare i due canali e veicolare all'occhio giusto il canale corrispondente. Purtroppo il costo ancora relativamente elevato (ma in discesa) dei proiettori e la necessità di spazio nel locale di proiezione ne possono limitare la diffusione.

Anche le tecniche basate su multiplex temporale e conseguente raddoppio della frequenza di quadro necessitano di monitor adatti: infatti i comuni televisori e i monitor al plasma e TFT non raggiungono la frequenza di quadro sufficiente[Nota 7], e i monitor per elaboratori elettronici (workstation), peraltro utilizzabili solo se basati su schermi CRT, oltre al costo non indifferente hanno una diversa colorimetria e falsano i colori. Si devono comunque utilizzare occhiali specifici, che possono essere di tipo passivo (con filtri polarizzatori o dicroici, rispettivamente adatti a sistemi con schermo ZScreen o alla tecnologia Infitec), o di tipo attivo (shutter), più pesanti e costosi.

Nota 6 - Si utilizzano polarizzazioni ortogonali di tipo circolare (destra/sinistra) o lineare (con filtri orientati tra loro a 90°). Nel campo della frequenza, ossia del colore, ci si riferisce al tradizionale anàglifo e al più raffinato Infitec commercializzato da Barco e basato su combinazioni di filtri dicroici.

Nota 7 - In genere i televisori non accettano più di 50Hz e gli altri apparati si fermano a circa 80Hz

5. Home cinema

Anche nel mercato dei video "da casa", cioè cassette e DVD, si inizia a parlare di stereoscopia.

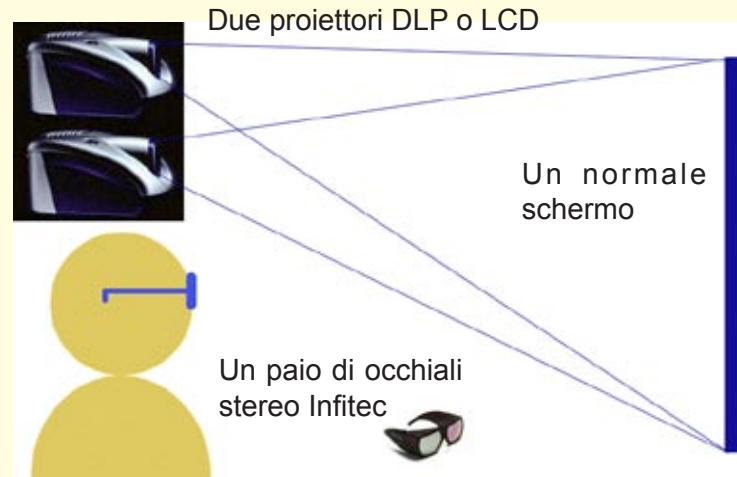
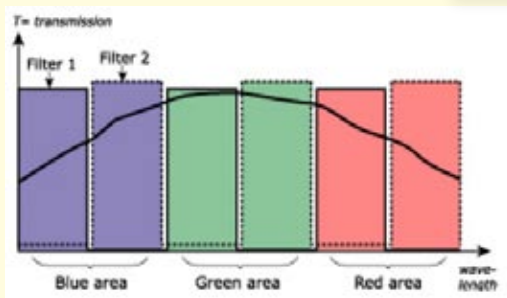
Delle varie applicazioni, questa è quella che, basandosi per motivi pratici ed economici su apparati standard di tipo consumer, subisce di più le loro limitazioni tecnologiche.

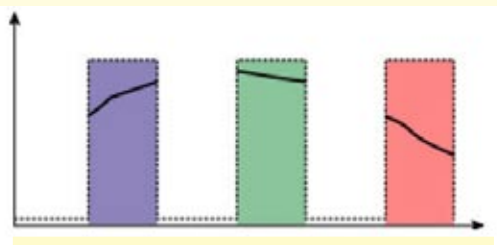
Le tecnologie stereoscopiche disponibili sono essenzialmente l'anàglifo, relativamente diffuso nel passato, ma che ora si tende a superare per via delle forti limitazioni che presenta, e la multiplazione temporale associata ad occhiali attivi di tipo shutter[Nota 8].

Nota 8 - Sono occhiali che al posto delle lenti hanno dei mini pannelli LCD, trasparenti o opachi a seconda della tensione loro applicata. Pilotati via cavo o tramite infrarossi da apposito apparato, vengono resi trasparenti o opachi alternativamente e in sincronia con il semiquadro visualizzato sul monitor, da cui la denominazione della tecnica *field-sequential.*

Fig. 7 - Sistema di visualizzazione stereoscopica per uso casalingo: sincronizzatore con trasmettitore IR e occhiali shutter "wireless"

IL SISTEMA INFITEC

Il sistema Infitec (acronimo di Interferenz Filter Technik) è il risultato dei un progetto di richeca della DaimlerChrysler AG di Ulm. Dal 2003 Infitec è un'azienda indipendente (www.infitec.net).

Schema del sistema di proiezione

Rappresentazione schematica della banda di frequenza del segnale completo

Segnale costituente il canale sinistro

Segnale costituente il canale destro

Si basa sulla separazione dei canali costituenti la coppia stereoscopica ottenuta nel campo delle frequenze (colori) con dei filtri ottici dicroici inclusi nel corpo di proiettori DLP o LCD costituenti una coppia di proiezione stereoscopica.

La proiezione si effettua secondo il solito schema basato su coppia di proiettori. Lo spettatore deve dotarsi di occhiali passivi (peraltro non particolarmente economici) equipaggiati da filtri ottici dicroici con caratteristiche uguali a quelli usati nei proiettori per separare i canali destro e sinistro.

Il sistema ha diversi pregi, tra i quali la possibilità di utilizzare comuni schermi bianchi anche diffusivi, presenta un'ottima separazione tra i canali senza effetto ghosting apprezzabile, permette movimento della testa, è esente da flickering e offre immagini molto luminose e ad alto contrasto.

Fig. 8 - Sistema di visualizzazione stereoscopica per uso casalingo. : analizzatore di scena bidimensionale con generazione di coppia stereoscopica, trasmettitore IR e occhiali shutter "wireless". Le dimensioni e il peso non differiscono dal semplice sincronizzatore di figura 7.

Si noti che per non raddoppiare la frequenza di quadro (solo alcuni apparati lo fanno, richiedendo i poco diffusi televisori a 100Hz, oppure monitor da personal computer) si adotta la tecnica chiamata field-sequential che consiste nell'inserire nei semiquadri pari il segnale relativo ad un canale della coppia stereo e nei semiquadri dispari il segnale relativo all'altro canale della coppia stereoscopica. In questo caso si rilevano prezzi decisamente economici (inferiori a 100 €), anche con pilotaggio ad infrarossi degli occhiali shutter, e una offerta di titoli su DVD scarsa in termini assoluti, ma interessante considerando l'ancora scarsa diffusione , e a prezzi ragionevoli (circa 20 €/DVD). Di contro c'è da attendersi un flicker notevole, dato che la frequenza di quadro del canale che raggiunge il singolo occhio è di 25 Hz[Nota 9].

Interessante è la presenza sul mercato di sistemi basati su tecnologia field-sequential che effettuano il deinterallacciamento, ottimizzando le prestazioni dei proiettori LCD e DLP e agli schermi al plasma e TFT, e di sistemi più complessi, ma a basso costo, che generano l'effetto stereoscopico tramite elaborazioni in tempo reale sul segnale bidimensionale[Nota 10].

Nota 9 - I commenti reperiti in Internet riguardo ai DVD 3D non sono molto positivi riguardo ai contenuti, ma non sono negativi rispetto all'effetto tridimensionale. Sembrerebbe pertanto che la tecnica adottata accontenti l'utenza dal punto di vista tecnico, mentre la scarsa quantità di titoli disponibili probabilmente non l'accontenta come contenuti.

Nota 10 - Si cita per esempio il "Virtual FX 3D converter" prodotto dalla eDimensional e venduto a circa 100 €

Conclusioni

La visione tridimensionale è la prossima frontiera che verrà affrontata dal mondo dello spettacolo per richiamare l'attenzione degli spettatori.

I vari settori mostrano un diverso grado di sviluppo, dall'affermato videogioco alla futuribile televisione, e hanno diverse ambizioni, in particolare relativamente ai risultati tecnici e all'appagamento dello spettatore, le maggiori delle quali espresse dall'industria cinematografica.

Nel mondo della televisione si riscontra un interesse nella stereoscopia piuttosto moderato, per ora limitato alla ricerca e alla normalizzazione, giustificato dagli investimenti in corso e a quelli previsti per il prossimo futuro relativi ai sistemi digitali e alla TV ad alta definizione. Tuttavia se le tecnologie di visione tridimensionale prenderanno piede negli altri settori dello spettacolo, in particolare nel cinema, è molto probabile che anche gli operatori televisivi saranno spinti ad adottare tali sistemi, risolvendo i problemi tecnici che ancora impediscono la fattibilità completa e a prezzi adeguati della TV stereoscopica.

Bibliografia

1. Riccardo Drago – Jolie è più bella in 3D – L'Espresso, 8 giugno 2006
2. Audrey Doyle - Spiare in stereoscopia – Computer Gazette, novembre 2003
3. Michael Starks – IMAX 3D – Terrific Format, Modest Results – 3D-TV, www.3dmagic.com/articles/imax.html
4. David Starkman – Four Million Housequests - www.3dgear.com/scsc/movies/imax.html
5. Andrew Weil – James Cameron on Battle Angel! - comingsoon.net, 10 gennaio 2005
6. Xeni Jardin - James Cameron's new 3D film: Aliens of the Deep – boingboing.com, 21 gennaio 2005
7. Bryant Fazer - Is it the most important part of the D-Cinema roll-out? - studiodaily.com, 10 Aprile 2006
8. TV of the future – theage.com.au, 28 maggio 2004
9. Roland Piquepaille - Holographic Movies For Your TV – 16 giugno 2005
10. Swanni Sez – 3-D TV On Display In Japan – TVPredictions.com, 21 aprile 2006
11. Courtney Ostaff, Jason D'Aprile – Japanese Public Broadcasting Envisions 3D Future – PhysOrg.com, 2005
12. Philips introduces new 42" 3D LCD display for a stunning 3D experience - http://www.business-sites.philips.com/3dsolutions/about/article-15169.html, 8 marzo 2006

13. Philips 3D intelligent display solution, based on WOWvx technology, enables exciting 3D viewing experiences for high-impact out-of-home advertising – http://www.business-sites.philips.com/3dsolutions/About/article-14922.html, 3 gennaio 2006

14. Philip Swann - 3-D TV In Two Years? Philips says it will launch the futuristic TV in the near future - http://www.tvpredictions.com/philips060906.htm, 9 giugno 2006

15. 3D Content Enabling Technology - http://www.business-sites.philips.com/3dsolutions/3dtechnology/3dcontentenablingtechnology/index.html

Si segnala che il numero di agosto 2004 di Elettronica e Telecomunicazioni contiene diversi articoli sul tema della televisione stereoscopica.

Progettazione di un sistema fuzzy per l'annotazione automatica di contenuti multimediali

ing. Maurizio Montagnuolo

ing. Alberto Messina
Rai - Centro Ricerche e Innovazione Tecnologica

L'articolo descrive parte dell'attività di ricerca svolta presso il Centro Ricerche e Innovazione Tecnologica della Rai nel campo della documentazione automatica degli archivi.
Questa ricerca è realizzata nell'ambito della collaborazione tra Centro Ricerche Rai e Università degli studi di Torino. La borsa di dottorato di ricerca dell'Ing. Montagnuolo è sponsorizzata da euriX s.r.l.

Sommario
Questo articolo presenta un framework per la caratterizzazione automatica di contenuti multimediali, basato su tecnologie di data mining, Semantic Web e logica fuzzy. L'approccio proposto è basato sul presupposto che i fruitori di contenuti multimediali utilizzino implicitamente un sistema di regole di inferenza fuzzy per la caratterizzazione semantica dei contenuti audiovisivi, e sull'asserzione che le regole utilizzate dai fruitori dei contenuti multimediali per inferire i concetti semantici di alto livello, nonché i concetti semantici stessi, non sono definite a priori su classificazioni statiche, bensì definite dinamicamente, sulla base degli schemi caratteristici derivati dalle caratteristiche spazio-temporali estratte dagli oggetti multimediali. L'uso delle ontologie [16] permette di esprimere i concetti semantici (oggetti, genere, eventi, azioni, ecc.) in modo formale, in maniera dipendente dal dominio e comprensibile dalle macchine, mentre l'uso della logica fuzzy permette di esprimere analiticamente il grado di incertezza insito nei modelli di classificazione.

1. Introduzione

Grazie al veloce sviluppo delle tecnologie informatiche, la disponibilità di contenuti digitali multimediali è in continua crescita[Nota 1]. In ambito professionale, i broadcaster televisivi pubblici europei rappresentano una miniera di contenuti di prima importanza in questo senso. In questo dominio applicativo in genere si rende necessario lo sviluppo di applicazioni che permettano l'organizzazione ed il recupero efficiente dell'informazione all'interno di database multimediali di ingenti dimensioni. Si consideri, ad esempio, il caso del reperimento di spezzoni audio/video soddisfacenti determinati criteri descrittivi, che rappresenta una delle operazioni più importanti, ma al contempo onerose, nell'ambito della produzione radiotelevisiva. Tipicamente, le operazioni di annotazione, indicizzazione e predisposizione alla ricerca degli archivi audiovisivi, costringono

Nota 1 - Si considerino ad esempio i motori di ricerca video YouTuBe (http://www.youtube.com/) e Google Video (http://video.google.it/)

i documentatori a diverse ore di lavoro manuale per l'annotazione degli archivi stessi, pertanto rendendo l'intero processo inefficiente e costoso. Dalla difficoltà di ottenere in tempi brevi una documentazione oggettiva e completa, derivano, infatti, gli alti costi dovuti alla produzione del materiale multimediale che gravano sull'industria dell'audiovisivo. Conseguentemente, si sta mettendo in opera, da parte dell'industria audiovisiva, un ingente sforzo nella ricerca di modelli produttivi che, sfruttando tutte le possibilità offerte dalle moderne tecnologie, consentano di ottenere notevoli tagli nei costi di produzione.

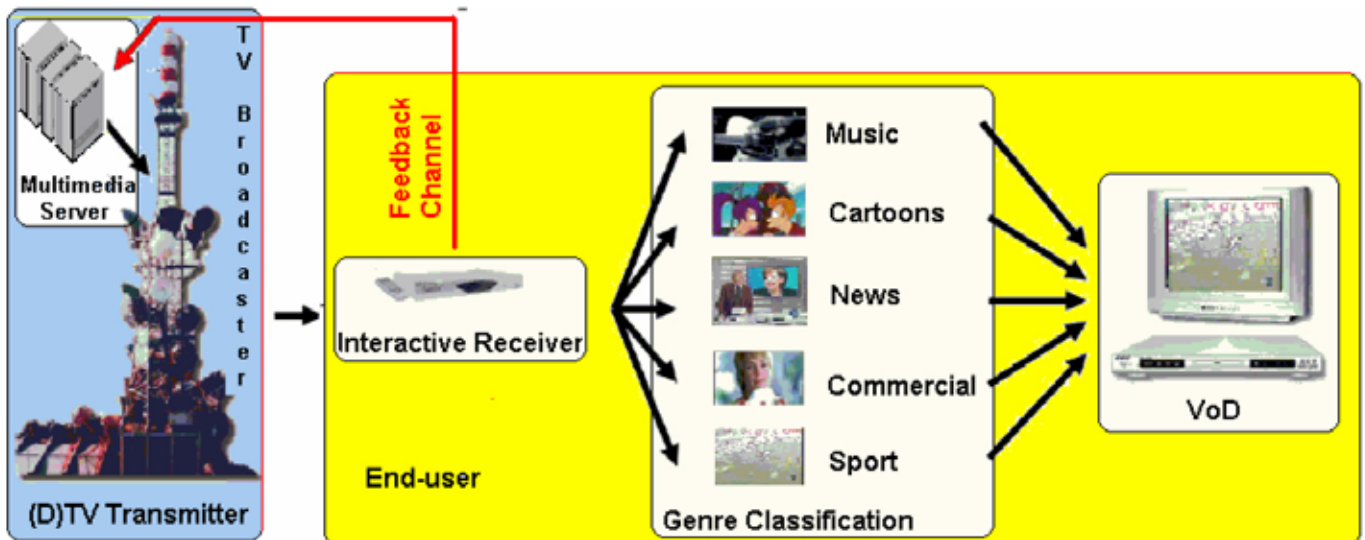
Piuttosto ovviamente, gli archivi multimediali sono una miniera di contenuti cui attingere per le nuove produzioni, e pertanto i metodi di reperimento ed accesso al materiale archiviato diventano sempre più importanti, nell'economia industriale e di gestione di un'impresa audiovisiva. Per la messa in opera di tali metodi, è necessario il coinvolgimento di ricercatori impegnati in diverse discipline dell'ingegneria, della fisica e dell'informatica, tra le quali possiamo annoverare: *Image Analysis, Pattern Recognition, Computer Vision, Database Multimediali, Reti Neurali e Software Engineering*. Nella comunità tecnico-scientifica internazionale, attualmente i maggiori sforzi sono concentrati nello studio di sistemi che, mediante una minima interazione con un utente umano, permettano di fornire una descrizione ed una classificazione di alto livello del contenuto di un oggetto multimediale. Infatti, considerando l'esempio degli archivi audiovisivi, la maggior parte degli attuali sistemi di *video retrieva*l basati sul contenuto (CBVRS – Content Based Video Retrieval Systems) utilizzano modelli matematici/statistici per la descrizione delle caratteristiche (*features*) di basso livello, quali il colore, la trama, la forma, l'audio ed il movimento. Questi modelli sono affetti da una grave problematica dovuta al fatto che un utente umano usa effettuare un'analisi del contenuto d'alto livello (semantico), mentre non ha in genere esplicita conoscenza della sua descrizione mediante modelli matematici. Di conseguenza, le metriche per valutare il grado di somiglianza tra oggetti multimediali sono spesso diverse se considerate dal punto di vista dell'utente o da quello della macchina. Per ovviare al suddetto problema è necessario prevedere degli strumenti automatici che permettano di determinare una correlazione tra le caratteristiche estratte di basso livello e le caratteristiche semantiche di alto livello. In quest'ambito si inserisce la nostra ricerca, avente lo scopo di analizzare lo stato dell'arte dei sistemi automatici di annotazione di oggetti multimediali, valutandone una possibile implementazione ed i relativi rischi, costi e benefici derivanti dal suo utilizzo nell'industria audiovisiva (per una prima analisi di queste problematiche vedi anche [15]).

2. Semantica in Ambito Radiotelevisivo

Nel dominio radiotelevisivo, in cui operiamo, l'annotazione semantica dei programmi televisivi esprime informazioni semantiche associate allo scopo del programma stesso (ad esempio *'intrattenimento' (entertainment)*, *'informazione' (information)*, *'comunicazione' (communication)*, etc.) al suo genere (ad esempio *'talk-show*', *'evento sportivo' (sport)*, *'telegiornale' (newcast)*, etc.), a concetti strutturali ed editoriali (ad esempio *'intervista' (interview)*, *'rapporto giornalistico' (report)*, *'conduzione' (anchorperson)*, etc. o ad eventi (ad esempio *'goal'*, *'esplosione'*, etc.).

Tra i diversi tipi di informazione sopraccitati, crediamo che la caratterizzazione automatica del genere rappresenti un punto di notevole interesse nella comprensione e classificazione dei contenuti multimediali. Tale tecnica, nota con il termine di *Teoria del genere*, permette di classificare i programmi televisivi in categorie rappresentative, caratterizzanti le intenzioni del regista, il ritmo e la struttura del programma, nonché ciò che uno spettatore si aspetta di osservare visionando il programma stesso.

A conferma dell'importanza che le tecniche di

Fig. 1 - Esempio di applicazione di Video-on-Demand (VoD) e televisione interattiva (iTV).
I contenuti multimediali sono classificati ed archiviati nei server del produttore televisivo.
L'utente può, per mezzo di un ricevitore interattivo, accedere ad uno specifico tipo di programma da lui desiderato e visionarlo sul proprio televisore.

classificazione automatica del genere (*Automatic Genre Classification* - AGC) rivestono nell'ambito dell'annotazione di archivi radiotelevisivi, sia i grandi broadcaster televisivi (tra cui Rai), sia le industrie di elettronica di consumo stanno prendendo parte attiva nella sperimentazione di tali tecniche. Infatti, tali strumenti permettono ai broadcaster di gestire in modo più efficiente i propri archivi multimediali, diminuendo di conseguenza i costi di produzione; inoltre permettono ai singoli utenti di creare liste di programmi personalizzate quando si utilizzano servizi di Video-on-Demand (VoD) e TV interattiva (iTV). In questo scenario, mostrato in figura 1, dato un flusso video digitale in ingresso, un ricevitore interattivo ne permetterà la segmentazione, indicizzazione, classificazione e memorizzazione in un hard disk in esso integrato. Quindi, un utente potrà accedere ad uno specifico tipo di programma da lui desiderato semplicemente navigando, tramite una pratica interfaccia grafica (Graphical User Interface - GUI) visualizzata sul proprio apparecchio televisivo, le cartelle situate nella memoria del ricevitore stesso e organizzate secondo generi e sottogeneri familiari all'utente.

Ma come è possibile costruire concretamente un sistema che attui in maniera efficiente questo tipo di scenari? Si cercherà di fornire una risposta esauriente a questo quesito nei paragrafi successivi.

3. Considerazioni sulla Classificazione Automatica del Genere

Fischer et al. [1] sono considerati i pionieri dello studio di metodologie per la classificazione automatica del genere di sequenze video. La loro ricerca, che ebbe inizio nel 1995, consisteva nella classificazione di cinque generi televisivi

(*news*, *commercials*, *cartoons*, *tennis* e *formula 1*) utilizzando una combinazione di caratteristiche audiovisive direttamente estraibili dalle sequenze video. Sulla base di tale ricerca seguirono altri approcci, distinguibili per la scelta dei descrittori e dei classificatori utilizzati [2, 3, 4, 5, 6, 7, 8]. Sebbene, a causa della non uniformità nella scelta del database di test, dei descrittori e del tipo di classificatore non sia possibile confrontare direttamente le prestazioni dei diversi metodi proposti, è possibile notare alcuni punti deboli comuni tra essi.

Per prima cosa, tutti i metodi proposti sono mirati alla classificazione statica in poche categorie predefinite, utilizzando un approccio ispirato dal paradigma "*è questo o quel tipo di genere*" (ad esempio, *"questo video è sport"*). In altri termini, pochi autori hanno considerato il caso in cui un contenuto multimediale sia caratterizzato da generi differenti. Consideriamo ad esempio la finale dei mondiali di calcio disputata in Spagna nel 1982. Questo programma era a contenuto sportivo nel 1982, ma ad oggi potrebbe essere classificato sia come *pubblicità*, sia come *documentario*. Ciò introduce una seconda limitazione, riguardante il fatto che il linguaggio televisivo è un linguaggio vivo che si evolve nel tempo in accordo con l'evoluzione linguistica, sociale e culturale della comunità in cui e per cui i programmi sono prodotti. Pertanto, confrontare programmi prodotti in epoche diverse o anche trasmessi su canali differenti, potrebbe rivelarsi un compito arduo, dal quale potrebbero inoltre derivare risultati inattesi od errati nel processo di annotazione del materiale audiovisivo. In termini astratti, si potrebbe dire che le caratteristiche di contenuto, strutturali e cognitive di un oggetto multimediale sono associabili a generi e sottogeneri in stretta dipendenza da una informazione contestuale, la quale è nativamente estranea ad essi, ma fondamentale per la definizione del genere stesso. Questa informazione contestuale è difficilmente rappresentabile in maniera analitica, ma di certo le condizioni storiche, le caratteristiche editoriali del canale di pubblicazione sono ad esempio elementi cardine in questo senso. Anche per questi motivi, gli autori di questa ricerca ritengono opportuno parlare di *caratterizzazione* più che di *classificazione*, intendendo per classificazione il complesso processo secondario che associa una caratterizzazione ad un concetto semantico per via di un contesto di fruizione.

Un ultimo problema che vogliamo menzionare riguarda il fatto che gli approcci sino ad ora proposti non hanno tenuto in considerazione l'informazione derivante dall'analisi spazio-temporale delle relazioni esistenti tra un intero programma e le sottostorie in esso contenute (ad esempio, un talk-show ed ogni singolo tema in esso discusso). Infatti, i database di test sono stati solitamente costruiti *ad-hoc* montando in successione piccole sequenze video aventi come unica caratteristica comune il fatto di appartenere, *a giudizio dell'autore della ricerca*, allo stesso genere; di conseguenza, le relazioni gerarchiche tra un intero programma ed i capitoli in esso contenuti non sono state prese in considerazione nell'analisi dei modelli di classificazione.

4. Progettazione di un Framework per l'Annotazione Automatica di Contenuti Multimediali

4.1 Requisiti

E' nostra opinione che nella progettazione di un sistema innovativo per l'annotazione automatica di contenuti multimediali devono essere soddisfatti i seguenti requisiti [17]:

- Prevedere un modello concettuale dei dati in grado di caratterizzare il dominio applicativo e definire le relazioni esistenti tra i diversi livelli di informazione insiti in un oggetto multimediale;

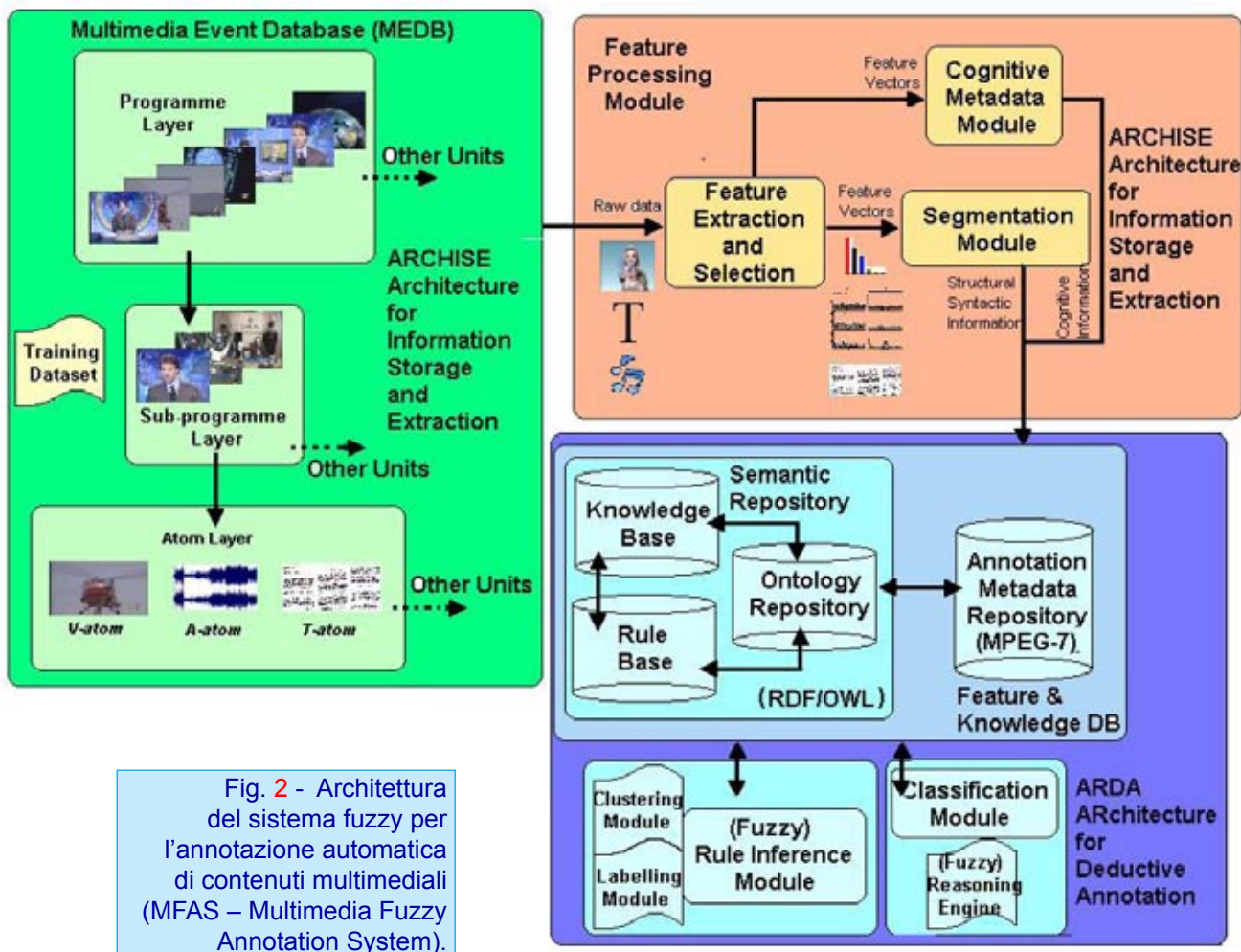



Fig. 2 - Architettura del sistema fuzzy per l'annotazione automatica di contenuti multimediali (MFAS – Multimedia Fuzzy Annotation System).

- Definire un modello di database per la rappresentazione efficiente del contenuto degli oggetti multimediali e delle relazioni esistenti tra oggetti differenti;
- Sviluppare un sistema che facendo uso delle tecnologie del Semantic Web, delle tecniche di Pattern Recognition e di regole di inferenza, permetta di associare le caratteristiche audiovisive di basso livello (automaticamente estraibili dagli oggetti multimediali) a concetti e classi semantiche definite secondo ontologie personalizzate in contesti di riferimento definiti.

4.2 Architettura del Sistema

La figura 2 mostra lo schema a blocchi di un sistema fuzzy per l'annotazione semantica di contenuti multimediali (MFAS - Multimedia Fuzzy Annotation System). Un sistema MFAS deve prevedere diversi sottosistemi cooperanti tra loro, ognuno dei quali designato all'analisi di uno specifico aspetto caratterizzante l'oggetto multimediale analizzato. Inoltre, a differenza delle immagini fisse come le fotografie, l'analisi di una sequenza multimediale deve tenere in considerazione aspetti di natura diversa, ognuno dei quali descrivente una specifica peculiarità

della sequenza stessa. Occorre ricordare che una sequenza video non può essere considerata come una successione di immagini indipendenti, poiché tutti i fotogrammi della sequenza sono tra loro strettamente correlati. Le transizioni tra un fotogramma ed il successivo, e le variazioni delle caratteristiche di colore, trama, e forma osservate analizzando gruppi di fotogrammi possono essere più importanti, ai nostri fini, dell'informazione contenuta in un singolo fotogramma. Inoltre, ogni oggetto ha caratteristiche peculiari che lo distinguono dagli altri, (velocità d'azione e di ripresa, linguaggio visivo, audio, ecc.) secondo la classe cui appartiene.

L'architettura proposta fa uso della logica fuzzy e tecniche di rule mining al fine di simulare il ragionamento umano nel compito di aggregare oggetti multimediali "simili tra loro". Riteniamo che questo compito elementare sia la base del processo di classificazione dei contenuti multimediali che avviene durante la fruizione dei contenuti stessi. L'uso di metodi di tipo fuzzy è preferibile ai metodi di classificazione classici (utilizzanti ad esempio reti neurali, reti bayesiane, ecc), poiché permette di tenere in considerazione la natura polivalente dei dati da analizzare. Infatti, poiché ogni singolo oggetto multimediale può appartenere ad una, nessuna, o più classi semantiche contemporaneamente, è più corretto fornire, per ogni oggetto, il suo grado di appartenenza ad ogni classe elementare definita, piuttosto che fornirne una classificazione netta del tipo "*questo oggetto appartiene alla classe A, piuttosto che alla classe B*". Con questa tecnica, l'obiettivo di costruire una classificazione semantica finale, passa attraverso un'analisi polivalente delle caratteristiche di contenuto.

Il sistema è composto da due moduli principali, brevemente descritti nei paragrafi seguenti:

i. Il modulo ARCHISE (**ARCH**itecture for **I**nformation **S**torage and **E**xtraction), utilizzato per estrarre l'informazione di basso-medio livello sul contenuto di un oggetto multimediale e per conservare l'insieme di oggetti di riferimento per la costruzione dei modelli di caratterizzazione (MEDB);

ii. Il modulo ARDA (**AR**chitecture for **D**eductive **A**nnotation), che è un sistema esperto che, simulando il ragionamento umano, permette di effettuare l'annotazione semantica degli oggetti multimediali.

4.2.1 Il Multimedia Event Database (MEDB)

Nel nostro scenario sperimentale, il Multimedia Event Database (MEDB) contiene circa 100 ore di programmi televisivi tratti dalle tre reti RAI e da altre emittenti locali. Ogni programma archiviato costituisce una *Programme Unit*, ovverosia un'entità multimediale fruibile singolarmente o appartenente ad un'altra unità. Al fine di mantenere la struttura coerente con le notazioni presenti in letteratura si è deciso di suddividere il materiale archiviato in sette classi semantiche, ciascuna di esse rappresentante uno specifico genere televisivo. Le classi individuate sono: *'Newscast'*, *'Commercials*', *'Cartoons'*, *'Football'*, *'Musics'*, *'Weather Forecasts*' e *'Talk Shows'*. Si noti che questa classificazione è usata per motivi meramente pratici di organizzazione del database, e non rappresenta la finalità della ricerca, che è invece, ricordiamo, quella di caratterizzare i contenuti multimediali attraverso un insieme di informazioni che servano ad associare in maniera dinamica, attraverso un processo dipendente dal contesto, gli oggetti multimediali stessi a classi semantiche significative dal punto di vista del fruitore.

4.2.2 Il Modulo di Processamento del Contenuto Audiovisivo

Il modulo di feature processing (FPM) è costituito da diversi sottosistemi atti all'estrazione ed al processamento del contenuto audiovisivo delle Programme Unit.

Il modulo di estrazione e selezione delle feature (FESM – Feature Extraction and Selection) estrae i descrittori audiovisivi di basso livello da ogni Programme Unit, e li rappresenta in conformità allo standard MPEG-7 [9].

Il modulo di segmentazione (SM) opera una segmentazione spazio-temporale della Programme Unit. Ogni sequenza è dapprima suddivisa in scene aventi caratteristiche visive omogenee utilizzando un algoritmo di similarità tra fotogrammi. Successivamente, un algoritmo di Image Clustering permette di raggruppare tra loro tutte le scene aventi grado di somiglianza superiore ad una certa soglia. Per ogni gruppo di scene simili (cluster) è fornito un identificatore univoco, il numero di scene ad esso appartenenti e i fotogrammi inziale, mediano e finale di ogni scena, nonché la posizione temporale di ciascun elemento facente parte del cluster.

Il modulo relativo all'estrazione dell'informazione cognitiva è costituito da motori di riconoscimento vocale (Speech-to-text) e facciale (Face recognition), provvedendo quindi ad un'estrazione di un tipo di informazione audiovisiva di medio livello semantico.

4.2.3 Il Modulo di Annotazione Semantica

ARDA (Architecture for Deductive Annotation) è un sistema intelligente mediante il quale è possibile inferire ed annotare automaticamente i concetti semantici intrinseci negli oggetti multimediali.

Tale procedura è realizzata utilizzando:

i. Un archivio semantico, contenente le ontologie descriventi i concetti da estrarre e rappresentare, la base di conoscenza del dominio applicativo e le regole d'inferenza;

ii. Un modulo fuzzy per il raggruppamento delle Programme Unit in accordo con le caratteristiche audiovisive, cognitive e strutturali da esse estratte;

iii. Un modulo di classificazione che è usato per assistere l'operatore nelle operazioni di classificazione e archiviazione del materiale audiovisivo.

4.2.3.1 L'uso delle Ontologie

Le ontologie [16] sono uno strumento utile nelle applicazioni di annotazione di contenuti multimediali poiché permettono di esprimere i concetti semantici (oggetti, genere, eventi, azioni, ecc.) in modo formale, dipendente dal dominio e comprensibile dalle macchine. Inoltre, è facile sviluppare ontologie personalizzate per la descrizione sia delle proprietà percettive e strutturali degli oggetti multimediali, sia delle relazioni esistenti tra tali proprietà. Infine, le ontologie sono strumenti scalabili e flessibili, che permettono dunque una facile integrazione e riuso delle stesse. In termini molto semplicistici un'ontologia è un dizionario di termini organizzato, nel quale è possibile definire relazioni gerarchiche tra termini, relazioni funzionali tra termini e condizioni che esprimono le proprietà delle relazioni stesse (p.es. transitività, relazione inversa).

Ovviamente, questa tecnologia presenta anche degli svantaggi. Il problema più grande consiste nel fatto che non esistono ancora ontologie standardizzate, e questo introduce di fatto delle ambiguità dovute alle diverse modalità con cui lo stesso concetto può essere rappresentato (si consideri ad esempio la lingua utilizzata per esprimere l'ontologia). Un secondo problema riguarda la dimensione delle ontologie stesse. Infatti, più un'ontologia è grande, più problematica risulta la sua gestione e riutilizzabilità. Per questo motivo sarebbe preferibile utilizzare un insieme di ontologie più piccole, ognuna di esse modellante un singolo aspetto del dominio applicativo o dell'informazione analizzata, piuttosto che un'unica grande ontologia. Ovviamente in questo secondo caso occorrerà arricchire la

base di conoscenza del sistema con le relazioni esistenti tra le diverse ontologie.

L'ultimo problema che vogliamo menzionare riguarda la scelta del tipo di linguaggio da utilizzare nella generazione dell'ontologia. Crediamo che la scelta migliore sia rappresentata dall'uso del linguaggio OWL (Web Ontology Language) [10]. Tale scelta è dettata dal fatto che l'OWL è un linguaggio standard raccomandato dal W3C. Inoltre, per la sua struttura e definizione, esso è più flessibile, intuitivo ed espressivo di altri linguaggi, quali ad esempio MPEG-7 XML, RDF e RDFS [11]. Infine, a tutt'oggi sono facilmente reperibili parser e motori di ragionamento automatico basati su tale linguaggio.

Nella nostra sperimentazione la tecnologia delle ontologie sarà primariamente adottata per la rappresentazione delle caratteristiche strutturali dei contenuti audiovisivi (p. es. numerosità dei cluster di immagini, loro mutua relazione) e per la derivazione automatica di proprietà secondarie godute dagli elementi strutturali e di conseguenza dalle Programme Unit che li contengono. Queste proprietà formeranno una base essenziale di informazioni per la fase di aggregazione e caratterizzazione delle Programme Unit.

4.2.3.2 L'uso della Logica Fuzzy

La logica comunemente utilizzata nei settori informatici e tecnologici è quella booleana, seguendo la quale ogni oggetto può essere membro di un unico insieme. Tale assioma è però in contrapposizione a quanto accade nel mondo reale, in cui le entità ad esso appartenenti non possono essere categoricamente riconosciute come appartenenti ad una o ad un'altra categoria in maniera esclusiva, ma possono appartenere ad entrambe con un certo grado di verità. Anche gli oggetti multimediali presentano tale proprietà, potendo appartenere, come già accennato precedentemente, ad una, nessuna, o più classi semantiche contemporaneamente. Consideriamo ad esempio un servizio sportivo all'interno del telegiornale; esso deve essere considerato nella categoria *'newscast'* o nella categoria *'sports'*? A nostro avviso può essere rappresentato da entrambe le categorie.

La logica fuzzy (dall'inglese "sfumato", "sfuocato") [12] è stata introdotta per formalizzare concetti del mondo reale che non possono essere categoricamente riconosciuti come veri o falsi, ma che possono avere un certo grado di verità. Dall'esempio precedente, appare evidente come, per le sue proprietà, la logica fuzzy, sia particolarmente efficace nelle applicazioni d'estrazione e interpretazione dell'informazione. Storicamente, la teoria sugli insiemi fuzzy, che pone le premesse della logica fuzzy, fu introdotta da Lotfi Zadeh[Nota 2] nel 1965, come strumento matematico per rappresentare i diversi livelli d'incertezza in campo linguistico. Inizialmente nacque come estensione della classica teoria degli insiemi.

In termini matematici tradizionali, un elemento dell'universo può appartenere o non appartenere ad un insieme. Cioè, l'appartenenza di un elemento è netta, rigida, (ovvero crisp), come espresso dall'equazione

$$\mu_A(x) = \begin{cases} 1 & se \quad x \in A \\ 0 & se \quad x \notin A \end{cases}$$

Un insieme fuzzy, viceversa, rappresenta una generalizzazione dell'insieme ordinario in cui non vengono solamente indicati gli elementi ad esso appartenenti, ma anche il grado di appartenenza di ciascuno di essi all'insieme stesso. Si perviene quindi alla seguente definizione di un insieme fuzzy:

data una collezione di oggetti x l'insieme A in X è l'insieme di coppie ordinate

$$A = \{(x, \mu_A(x)) : x \in X \}$$

Nota 2 - http://en.wikipedia.org/wiki/Lotfi_Zadeh

in cui $\mu_A(x) \in [0,1]$ indica il grado di appartenenza dell'oggetto x all'insieme A.

Da quanto enunciato, il lettore potrebbe confondere la logica fuzzy con la teoria della probabilità. Questa confusione può essere chiarita, se pensiamo che la logica fuzzy tratta eventi deterministici (ovvero fatti tangibili, misurabili), mentre la teoria della probabilità riguarda la verosimiglianza di eventi non deterministici (ovvero stocastici). Pertanto, la logica fuzzy esprime l'incertezza riscontrabile nella definizione di un concetto, fornendo il grado di similitudine di oggetti nei confronti di particolari proprietà intrinseche dell'oggetto, mentre la teoria della probabilità esprime l'incertezza dell'occorrenza di fenomeni, fornendo la probabilità del verificarsi di un evento sulla base delle sue occorrenze.

Nel nostro sistema la logica fuzzy è utilizzata per effettuare un raggruppamento di Programme Unit aventi caratteristiche strutturali, cognitive o audiovisive similari; ciò viene effettuato applicando l'algoritmo Fuzzy C-Means (FCM) [13] monodimensionale su ogni vettore n-dimensionale contenente i valori numerici dei descrittori estratti da ogni Programme Unit. L'algoritmo FCM è una procedura iterativa in cui ogni cluster c_i (*i = 1, 2, …, C*) è considerato come un insieme fuzzy. Ogni elemento dello spazio considerato x_j (*j = 1, 2, …, P*) può pertanto appartenere simultaneamente a più cluster con differenti indici di appartenenza u_{ij} (*i = 1, 2, …,* C; *j = 1, 2, .., P*), variabili ad ogni iterazione dell'algoritmo. L'agoritmo opera come segue:

1. *INIZIALIZZAZIONE:*

a. *Scegliere il numero di cluster C per il numero P di elementi da raggruppare;*

b. *Scegliere il valore del livello di ambiguità (dall'inglese fuzzyness) F (tipicamente F = 2). Tale parametro è utilizzato per controllare il livello di ambiguità tra i diversi cluster.*

Per F = 0 i cluster sono di tipo classico, in cui ogni elemento può appartenere ad un solo cluster (algoritmo k-means), mentre per F >> 0 il livello di ambiguità tra i cluster aumenta;

c. *Scegliere la funzione di distanza, utilizzata per determinare il grado di appartenenza di un elemento ai cluster;*

d. Scegliere un criterio di distanza tra le matrici dei gradi di appartenenza, stimate per iterazioni successive $\|U(t+1)-U(t)\|$;

e. Scegliere la soglia δ che determina la condizione di termine dell'algoritmo;

f. Inizializzare la matrice U(t=0)

2. *AGGIORNAMENTO DEL BARICENTRO DI OGNI CLUSTER:*

$$\forall i = 1,2,\ldots,C \quad b_i(t+1) = \frac{\sum_{j=1}^{P} [u_{ij}(t)]^F X^j}{\sum_{j=1}^{P} [u_{ij}(t)]^F}$$

3. *AGGIORNAMENTO DEI GRADI DI APPARTENENZA:*

$$\forall i = 1,2,\ldots,C \;\; e \;\; \forall j = 1,2,\ldots,P$$
$$u_{ij}(t+1) = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|X^j - b_i(t+1)\|}{\|X^j - b_k(t+1)\|} \right)^{\frac{2}{F-1}}}$$

4. *CONDIZIONE DI TERMINE:*

if $\|U(t+1)-U(t)\| \leq \delta$

termina algoritmo

else

t = t+1

goto 2

L'algoritmo FCM ha il vantaggio di ottenere un buon compromesso tra l'efficacia della aggregazione (clustering) ed il costo computazionale. Dall'altra parte, la sua efficienza dipende fortemente dalla scelta iniziale della matrice U(t=0).

Dopo aver applicato l'algoritmo FCM ogni Programme Unit sarà identificata, per ogni data caratteristica da un vettore contenente i gradi di appartenenza ad ogni cluster; mediante un'operazione di fuzzificazione potrebbe essere pertanto possibile associare i centroidi di ogni cluster ad un'etichetta linguistica definita nell'ontologia, producendo una mappatura tra i descrittori di basso livello ed i concetti semantici di livello superiore.

4.2.3.3 L'uso di Tecniche di Data Mining

Il modulo di inferenza delle regole (RIM) è utilizzato per identificare le relazioni più frequenti esistenti tra i descrittori e di concetti. Tali relazioni sono espresse sotto forma di regole di associazione $R = \{r_1, r_2, \ldots, r_m\}$, memorizzate in un apposito database (RB – Rule Base). Ogni regola associa $r : F^{m,n} \rightarrow C$ una sequenza di descrittori di basso livello f appartenente all'insieme dei F dei descrittori, alle classi semantiche di alto livello, definite nell'insieme dei concetti C ed espressi da apposite ontologie. Le regole sono generate applicando tecniche di data mining [14] ad ogni transazione $T = \{t_1, t_2, \ldots, t_m\}$. Formalmente, $t_i = \overline{f} \cup \overline{c}$, in cui f è una possibile combinazione dei vettori delle feature e c una possibile combinazione dei concetti semantici rappresentati nell'ontologia. Queste tecniche saranno applicate a valle dell'applicazione di FCM sugli insiemi fuzzy generati.

5. Conclusioni

L'articolo ha esaminato e descritto a grandi linee l'architettura di un sistema per la caratterizzazione e l'annotazione automatica di contenuti multimediali, facente uso della logica fuzzy, tecnologie del semantic web e tecniche di data mining. Crediamo che, utilizzando tecniche di analisi del contenuto audiovisivo multimodali, in unione con le tecnologie del Web semantico e sistemi intelligenti sia possibile ottenere un sistema flessibile ed affidabile, in grado di avanzare lo stato nell'arte nella minimizzazione del gap semantico esistente tra la rappresentazione delle caratteristiche audiovisive di basso livello e la deduzione di concetti semantici di livello più alto.
Questo sistema è attualmente oggetto di studio e sperimentazione presso il Centro Ricerche e Innovazione Tecnologica della RAI, attraverso la collaborazione di alcuni ricercatori RAI e di personale dell'Università degli studi di Torino che fruisce della sponsorizzazione della borsa di dottorato di ricerca da parte di euriX s.r.l.

Bibliografia

1. Fischer S., Lienhart R. and Effelsberg W. (1995), Automatic recognition of film genres, in ACM Multimedia 1995, San Francisco, USA, pp. 295-304.

2. Truong B.A. and Dorai C. (2000), Automatic genre identification for content-based video categorization, Proceedings of the International Conference on Pattern Recognition, Barcelona, Spain, Vol. 4, pp. 230-233.

3. Xu L.Q. and Li Y. (2003), Video classification using spatial-temporal features and PCA, Proceedings of the IEEE International Conference on Multimedia and Expo (ICME2003), Baltimore, MD, USA.

4. Glasberg R., Elazouzi K. and Sikora T. (2005), Cartoon-Recognition using Visual-Descriptors and a Multilayer Perceptron, WIAMIS, Montreux, May 28-31.

5. Dimitrova N, Agnihotri L. and Wei G. (2000), Video classification based on HMM using text and faces, In European Signal Processing Conference, Tampere, Finland.

6. Liu Z., Huang J. and Wang Y. (1998), Classification of TV programs based on audio information using hidden Markov Model, Proc. of the IEEE Signal Processing Society Workshop on Multimedia Signal Processing.

7. Roach M.J., Mason J.S.D. and Pawlewski M. (2001),Video genre classification using dynamics, In ICASSP'2001.

8. Dinh P.Q., Dorai C. and Venkatesh S. (2002), Video genre categorization using audio wavelet coefficients, In ACCV 2002.

9. ISO/IEC 15398 Multimedia Content Description Interface, 2001.

10. WWW Consortium (W3C) (2004). Web Ontology Language (OWL), http://www.w3.org/2004/OWL/

11. WWW Consortium (W3C) (1999). Resource Description Framework (RDF), http://www.w3.org/RDF/

12. Timothy J. Ross (2004), Fuzzy Logic with Engineering Applications, 2nd Edition, Wiley

13. Frank Höppner, Frank Klawonn, Rudolf Kruse, Thomas Runkler (1999), Fuzzy Cluster Analysis, Wiley

14. Ian H. Witten, Eibe Frank (2005), Data Mining: Practical Machine Learning Tools and Techniques (Second Edition), Morgan Kaufmann.

15. A. Messina, D. Airola Gnota, "Automatic Archive Documentation based on Content Analysis", IBC 2005, Amsterdam 11 September 2005.

16. WWW Consortium (W3C) (2001), Semantic Web, http://www.w3.org/2001/sw/

17. M. Montagnuolo and A. Messina, "Multimedia Knowledge Representation for Automatic Annotation of Broadcast TV Archives", Proceedings of the 4th Special Workshop on Multimedia Semantics (WMS06), Chania, Crete, Greece, June 19-21, 2006, pp. 80-94.